Geophysical Journal International

Geophys. J. Int. (2020) **220,** 1078–1094
Advance Access publication 2019 November 06
GJI General Geophysical Methods

doi: 10.1093/gji/ggz503

Optimizing geophysical muon radiography using information theory

C.J. Benton,[1] C.N. Mitchell,[1] M. Coleman,[2] S.M. Paling,[3] D.L. Lincoln,[4] L. Thompson,[5] S.J. Clark[6] and J.G. Gluyas[6]

[1]*Department of Electronic and Electrical Engineering, University of Bath, Claverton Down, Bath BA27AY, UK. E-mail: cjb30@bath.ac.uk*
[2]*Jet Propulsion Laboratory, California Institute of Technology, M/S* 306-431, 4800 *Oak Grove Dr., Pasadena, CA* 91109, *USA*
[3]*STFC Boulby Underground Science Facility, Boulby Mine, Redcar-and-Cleveland, TS*134*UZ, UK*
[4]*Atkins, Chadwick House, Birchwood Park, Warrington, WA*36*AE, UK*
[5]*Department of Physics and Astronomy, University of Sheffield, Hicks Building, Hounsfield Road, Sheffield S*37*RH, UK*
[6]*Department of Earth Sciences, Durham University, Science Labs, Durham DH*13*LE, UK*

Accepted 2019 November 5. Received 2019 May 8; in original form 2019 October 23

SUMMARY

Cosmic ray muons are highly penetrating, with some reaching several kilometres into solid rock. Consequently, muon detectors have been used to probe the interiors of large geological structures, by observing how the muon flux varies with direction of arrival. There is an increasing need to discriminate between materials differing only slightly in bulk density. A particularly demanding application is in monitoring underground reservoirs used for CO_2 capture and storage, where bulk density changes of approximately 1 per cent are anticipated. Muon arrival is a random process, and it is the underlying expectation values, not the actual muon counts, which provide information on the physical parameters of the system. It is therefore necessary to distinguish between differences in muon counts due to real geological features, and those arising from random error. This is crucial in the low-contrast case, where the method can reach the information theoretic limit of what a data source can reveal, even in principle. To this end, methods to analyse information availability in low-contrast muon radiography have been developed, as have means to optimally interpret the available data, both for radiography and for tomography. This includes a method for calculating expectation values of muon flux for a given geological model directly, complementing existing Monte Carlo techniques. A case study, using a model of carbon capture is presented. It is shown that the new data analysis techniques have the potential to approximately double the effective sensitivity of the detectors.

Key words: Inverse theory; Probability distributions; Tomography.

1 INTRODUCTION

Cosmic ray muons are highly penetrating, with some reaching kilometres into solid rock before attenuation. Consequently, muon detectors have been used to probe the interiors of large geological structures such as volcanoes, by observing how the muon flux varies with direction (Nagamine *et al.* 1995; Tanaka *et al.* 2007; Lesparre *et al.* 2012; Carbone *et al.* 2013). The technique has also been applied to large artificial structures, such as Egyptian and Mesoamerican pyramids (Burkhard *et al.* 1970; Alfaro *et al.* 2003; Morishima *et al.* 2017), nuclear reactors (Perry *et al.* 2013; Morris *et al.* 2014) and water towers (Jourde *et al.* 2013).

A new application for muon radiography is monitoring underground reservoirs used for CO_2 storage (Kudryavtsev *et al.* 2011, 2012; Jiang *et al.* 2013). Carbon capture and storage schemes offer one of the few ways for reducing emissions of CO_2 to within safe limits, while prolonging the availability of fossil fuel based energy systems until a more sustainable infrastructure is developed. Once captured from power plants and industrial point sources, the CO_2 is compressed and transported to a suitable burial site. At this site, supercritical CO_2 fluid is injected through a well into a suitably porous and permeable geological formation, in which it displaces the fluid (e.g. brine) in the pore space. The difference in density (within the deep geological environments of interest) between supercritical CO_2 ($\lesssim$0.8 g cc^{-1}) and the preexisting pore fluid ($\gtrsim$1.0 g cc^{-1} for brine) is sufficient to allow the injection and emplacement processes to be monitored.

Prospective storage sites are usually deep (of the order 1 km) saline aquifers. Substantial study has gone into understanding the storage capacity, and the coupled physical–chemical processes of injecting CO_2 into geological formations (Mathias *et al.* 2013a,b). Until now, the monitoring of offshore deep storage sites has been largely based on the oil industry's standard practice of 3-D seismic surveys. The change in the response of a seismic signal can

be interpreted in terms of changes in temperature, pressure or (as caused by injection) fluid composition. However the signal is also affected by the prevailing sea conditions and errors in repositioning the detectors from survey to survey. These surveys are also very expensive, and are usually done infrequently. With muon radiography, however, the signal is both passive and continuous (Nakamura *et al.* 2010). It therefore has the potential to complement existing monitoring technologies and save long-term costs.

Conceptually, the muon radiography process works (Klinger *et al.* 2015; Gluyas *et al.* 2018), but there are various technological problems to overcome before it can become an industry standard method. There are also problems to be overcome when analysing muon data. One such problem is that muon tomography is an inherently random process, due to the cosmic rays arriving randomly, and the dynamics of muons in bulk materials being so complex as to be effectively random, with the material properties only affecting the probability distributions. This imposes an upper bound to the data acquisition rate, as it is necessary to distinguish between differences in muon counts due to real changes in the geology and those simply arising from random chance. The monitoring of carbon capture schemes is a particularly demanding application in this respect, as the expected changes in the muon arrival rate are in the order of 1 per cent. Therefore, rigorous analysis of the available data is required.

In Section 2, muon radiography is analysed using information theory, to determine the upper limit to the information that can be gathered, and to develop a technique for optimally extracting this data. This technique requires the expectation values of the muon flux, and hence in Section 3 a muon propagation model based directly on these probabilities is described. In both of these sections, the analysis is performed without reference to CO_2 storage, and hence remains valid for very low contrast muon radiography in general. In Section 4, a case study based on CO_2 storage under the North Sea is presented, and the benefit provided by the improved data analysis techniques is assessed. The key findings are summarized in Section 5.

2 APPLYING INFORMATION THEORY

Information theory can be used to explore the absolute upper limits to what can be inferred, even when all other parts of the system are perfect. The upper limits have been previously explored in the context of creating geological images with a specified contrast (Lesparre *et al.* 2010). This paper pushes further, into the regime where such an image would lack any discernible features, but where by rigorous analysis it can be possible to extract a few meaningful data bits. To achieve this, it is assumed that a model of the geology exists, and that the muon observations are being used to determine the values of certain unknown geological quantities.

In Section 2.1, muon radiography will be analysed in its most abstract sense, as a data channel which transfers information from the geology to the muon detectors. This provides the upper limit to the information that muon flux measurements can provide. In Sections 2.2 and 2.3, an optimal method for discriminating between hypotheses for unknown geological quantities is described. In Section 2.4, the results are converted into a form which takes into account the geometry of the detectors.

2.1 Analysis using Poisson communication theory

The upper limit to the available information can be calculated using a branch of information theory known as Poisson communication theory. This is a method largely developed to analyse light detectors operating on a photon by photon basis. An extensive overview of the subject is given by Verdu (1990).

The analysis starts with the assumption that there are a set of measurements, each consisting of a muon count. These may either be the total number of muons observed by a particular detector, or a subtotal corresponding to a subset of incoming directions.

It is also assumed that the muons arrive at the detector independently, in which case the counts will follow a Poisson distribution. This relies upon the fact that although multiple muons can be generated from the same cosmic ray, the chance of more than one of these reaching a given underground detector is negligible. While exceptional events can occur (Avati *et al.* 2003), they can be identified by multimuon arrival within an exceptionally short period of time.

If the counts are indeed Poisson distributed, then for the mth measurement, the probability P of observing k_m muons, given a particular geological hypothesis, is given by

$$\mathrm{P}(k_\mathrm{m}) = \frac{(\lambda_m)^{k_\mathrm{m}} \exp(-\lambda_m)}{k_\mathrm{m}!}, \tag{1}$$

where λ_m is the expectation value for that measurement and hypothesis. The expectation values are mostly dependent on observation time, and prior knowledge of the geology, whereas the quantities to be measured will only affect λ_m slightly. It is therefore useful to separate these out, by splitting λ_m as

$$\lambda_m = t\mu_\mathrm{m}, \tag{2}$$

where t is observation time and μ_m are mean muon arrival rates, and then splitting this further as

$$\mu_\mathrm{m} = \tilde{\mu}_\mathrm{m}(1+\eta_m), \tag{3}$$

where $\tilde{\mu}_\mathrm{m}$ are typical expectation values of the muon count rates. Depending on how exactly the method is applied, this may be a null hypothesis, or a starting value, or both. Deviations from this typical rate, corresponding to the geological quantities being measured, are given by η_m.

It is assumed that the values of $\tilde{\mu}_\mathrm{m}$ are known perfectly, with all uncertainty confined to η_m. In practice, values of $\tilde{\mu}_\mathrm{m}$ will themselves be subject to error, but so long as these errors are much smaller than the anticipated ranges of η_m, the assumption remains valid.

When η_m is dependent on more than one of the geological quantities to be measured, it can be split to first order as

$$\eta_m = \sum_n^N \zeta_{nm}\chi_n, \tag{4}$$

where χ_n is the value of the nth geological quantity (out of N), and the coefficients ζ_{nm} specify the extent to which the nth quantity affects the mth measurement. Strictly speaking, the accumulated are more likely to combine via multiplication, but as $\prod_n(1+\zeta_{nm}\chi_n) \approx 1+\sum_n\zeta_{nm}\chi_n$ this will make little difference.

The phrase 'geological quantity' is used here to describe any parameter of the model that is (to leading order) linearly related to a change in expected muon detection rates, and that is to be measured by the radiography. Examples may include material densities, thicknesses of geological layers, and (as done in Section 4) the extent of a plume of injected fluid.

It follows from eqs (3) and (4), that given a model of how μ_m depends on χ_n, the values of ζ_{nm} can be extracted as

$$\zeta_{nm} = \left[\frac{1}{\mu_\mathrm{m}}\frac{\partial\mu_\mathrm{m}}{\partial\chi_n}\right]_{\chi_{1\cdots N}=0}. \tag{5}$$

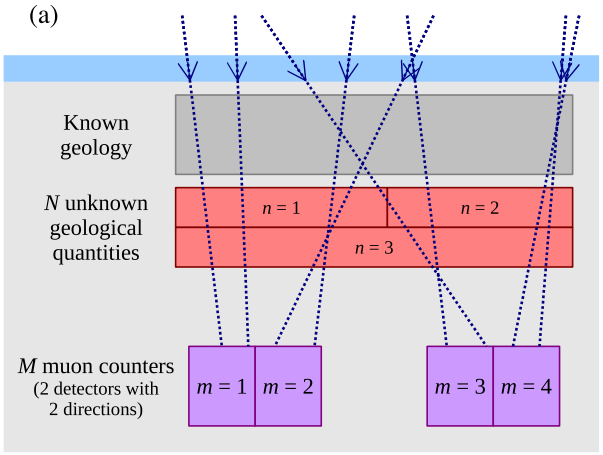

Figure 1. (a) Schematic diagram of geological imaging. Muons pass through regions of both known and unknown composition, and their direction of arrival logged by detectors. For the purpose of illustration, two detectors, splitting the incoming muons into two direction-of-arrival bins, are shown, giving a total of $M = 4$ muon counts. The unknown geology is represented by $N = 3$ quantities, which in this case correspond to the densities of different regions. For clarity, the known and unknown geological quantities are shown as corresponding to different locations. In general, the unknown quantities are modifications to known typical values at the same location. (b) The above, represented as an abstract form of data transfer, as described in the text.

A schematic diagram of the geology parametrization is given in Fig. 1.

In order to apply Poisson communication theory to muon radiography, the relevant problem is that of the Multi Input Multi Output (MIMO) channel with a rate amplitude constraint and very high dark noise. In the context of optics, this can be described as follows: The inputs are photon emitters, where the mean rate of photon emission can be controlled, but the release of individual photons cannot. The release of each photon is assumed to be independent of any other photon, and so the number of photons released over any time interval is Poisson distributed. For the nth emitter, the mean rate x_n can be varied over the range

$$0 \leq x_n \leq A_n, \tag{6}$$

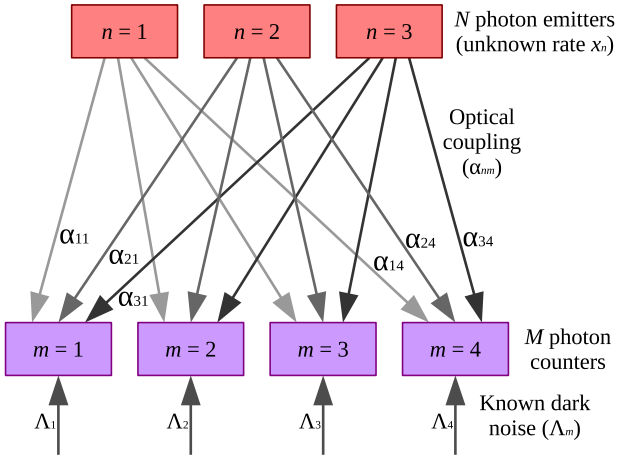

Figure 2. Schematic diagram of the optical communication system described in the text. The structure is the same as the geological imaging scheme shown in Fig. 1(b), allowing an analogy to be drawn.

where A_n is the rate amplitude constraint. Similarly, the outputs are photon detectors, which report photon detections at a rate

$$\mu_{\rm m} = \Lambda_m + \sum_n^N \alpha_{nm} x_n, \tag{7}$$

where α_{nm} is the efficiency of light transfer from the nth emitter to the mth detector. The detectors are assumed to have a dark noise, so that they will falsely register photon hits at a rate of Λ_m. This is assumed to be a Poisson process, and independent of real photon hits, so that it merely adds to the mean detected rate, as is done in eq. (7). The dark noise is further assumed to be very high, so that $\Lambda_m \gg x_n$. A schematic diagram of this system is shown in Fig. 2.

No paper seems to deal with the precise combination of requirements described here. In the Single Input Single Output (SISO) case, for very high noise, Davis (1980), extending on the work of Kabanov (1978), gives the channel capacity as

$$C_{\rm SISO} = \frac{(\alpha A)^2}{8\Lambda}, \tag{8}$$

where α and A are the single element equivalents of α_{nm} and A_n. This is Davis's eq. (8) a, where in his notation, αA is represented by c, and Λ is represented by λ_0. The rate is given in nats per unit time, where 1 nat is the equivalent of $\ln(2)$ data bits. Haas & Shapiro (2002) consider the MIMO case, but do not make the same investigation as Davis into very high noise limit. As is shown in Appendix A, the techniques can be combined to give

$$C_{\rm MIMO} = \sum_m^M \frac{1}{8\Lambda_m} \left(\sum_n^N \alpha_{nm} A_n \right)^2 \tag{9}$$

which as expected, reduces to eq. (8) in the SISO case.

It is now necessary to draw an analogy between this and the muon radiography problem. The first assumption is that muon counts are analogous to photon counts. Therefore, linking the rates $\mu_{\rm m}$ in eqs (3) and (7) gives

$$\tilde{\mu}_{\rm m} + \sum_n^N \tilde{\mu}_{\rm m} \zeta_{nm} \chi_n = \Lambda_m + \sum_n^N \alpha_{nm} x_n. \tag{10}$$

There is a slight complication, in that Λ_m is a lower bound, whereas $\tilde{\mu}_{\rm m}$ may not be. It is therefore useful to define $\eta_m^{\rm min}$, as the minimum possible values of η_m. If the values of χ_n are bounded by the closed

interval $\left[\chi_n^{\min}, \chi_n^{\max}\right]$, then by eq. (4) this minimum is given by

$$\eta_m^{\min} = \sum_n^N \min\left(\zeta_{nm}\chi_n^{\min}, \zeta_{nm}\chi_n^{\max}\right) \equiv \sum_n^N \zeta_{nm}\chi_n^- \tag{11}$$

where χ_n^- is shorthand for the value of $\chi_n^{\min}$ or $\chi_n^{\max}$ which provides the minimum term. Adding $\eta_m^{\min}$ to the first term of eq. (11) and then subtracting it from the second term gives

$$\tilde{\mu}_{\mathrm{m}}\left(1+\eta_m^{\min}\right) + \sum_n^N \tilde{\mu}_{\mathrm{m}}\zeta_{nm}\left(\chi_n - \chi_n^-\right) = \Lambda_m + \sum_n^N \alpha_{nm}x_n. \tag{12}$$

Both sides of the equation now consist of non-negative values being added to lower bounds. This allows the photon counting and muon counting terms to be equated as

$$\Lambda_m = \tilde{\mu}_{\mathrm{m}}\left(1+\eta_m^{\min}\right) \tag{13}$$

$$\alpha_{nm}x_n = \tilde{\mu}_{\mathrm{m}}\zeta_{nm}\left(\chi_n - \chi_n^-\right) \tag{14}$$

The second assumption is that the range of possible expectation values for the muon count rate is equivalent to the range of possible expectation values for the photon count rate. In other words, when the expectation value of the muon count rate is as low as believed possible (for a given geological model), this is equivalent to dark noise, and similarly, when the expectation value is as high as believed possible, this is equivalent to the maximum photon intensity case, where $x_n = A_n$. The former limit is given by eq. (13), whereas the latter is given by

$$\alpha_{nm}A_n = \tilde{\mu}_{\mathrm{m}}\zeta_{nm}\left(\chi_n^+ - \chi_n^-\right) \tag{15}$$

where χ_n^+ are the values of χ_n which maximize η_m. It is convenient to rewrite this as

$$\alpha_{nm}A_n = \tilde{\mu}_m\left|\zeta_{nm}\right|\Delta\chi_n \tag{16}$$

where

$$\Delta\chi_n \equiv \left|\chi_n^+ - \chi_n^-\right| \tag{17}$$

The reason for the modulus signs is that $\chi_n^+ - \chi_n^-$ is not necessarily positive, as χ_n^+ and χ_n^- correspond to maximizations and minimizations of η_m and not to the individual values of χ_n. If this difference happens to be negative, then ζ_{nm} will also be negative, and so eq. (16) remains valid.

Making the equivalences, using eqs (13) and (16), gives the data rate as

$$C = \frac{1}{8}\sum_m^M \tilde{\mu}_m\left(\sum_n^N \left|\zeta_{nm}\right|\Delta\chi_n\right)^2 \tag{18}$$

where due to $\eta_m^{\min}$ being small, the approximation $1+\eta_m^{\min} \approx 1$ has been made. Converting from nats to bits gives the binary data rate C_{b} as

$$C_{\mathrm{b}} = \frac{\ln(2)}{8}\sum_m^M \tilde{\mu}_m\left(\sum_n^N \left|\zeta_{nm}\right|\Delta\chi_n\right)^2 \tag{19}$$

This equation shows how much data can be inferred, but only as a total, and only in an abstract sense. The task of obtaining geological information from the muon counts still remains.

2.2 Orthogonal combinations of geological quantities

Information can be obtained from muon detector data by combining it into test statistics, and then using these to discriminate between different hypotheses for the geological quantities of interest. Before doing so, it is necessary to construct combinations of geological quantities such that the test statistics are independent. This can be done by switching to an alternative basis, consisting of linear combinations of geological quantities. For N such quantities, the ith linear combination χ_i' is defined by

$$\chi_i' = \sum_n^N X_{in}\chi_n, \tag{20}$$

where X is an $N \times N$ matrix. Writing eq. (4) in this basis gives

$$\eta_m = \sum_i^N \zeta_{im}'\chi_i', \tag{21}$$

where ζ_{im}' is defined in terms of the matrix inverse of X as

$$\zeta_{im}' = \sum_n^N X_{in}^{-1}\zeta_{nm}. \tag{22}$$

It will become apparent in Section 2.3 that the optimal test statistic requires ζ_n' to be orthogonal, such that

$$\sum_m^M \zeta_{im}'\zeta_{jm}' = \xi_i\delta_{ij}, \tag{23}$$

where δ_{ij} is the Kronecker delta, and where ξ_i are positive coefficients. It is shown in Appendix B that so long as there are N linearly independent sets of ζ_{nm}, this condition can be satisfied by defining the matrix T as

$$T_{in} = \sum_m \zeta_{nm}\zeta_{im} \tag{24}$$

in which case the ith row of X is the ith normalized eigenvector of T, with ξ_i being the corresponding eigenvalue. It is convenient to order the eigenvectors, and to choose their sign, so that the trace of X is maximized. This ordering results in X (which is already orthogonal) having unit determinant, in which case it becomes a rotation matrix. When there are two or three geological quantities, the trace is related to the rotation angle by

$$\theta_2 = \cos^{-1}\left(\frac{1}{2}\left(\mathrm{Tr}\,X_2\right)\right) \tag{25}$$

$$\theta_3 = \cos^{-1}\left(\frac{1}{2}\left(\mathrm{Tr}\,X_3 - 1\right)\right) \tag{26}$$

and therefore maximization of the trace amounts to minimization of the rotation angle. Higher dimensional analogies are more complicated, as general rotations can no longer be described by a single angle.

In all cases, however, the trace maximization can be thought of as preserving some degree of ordering. Where possible, it matches the elements of χ' to the elements of χ they depend on most. In the case where a particular geological quantity can already be observed independently, that is where muons passing through regions affected by that quantity will always miss regions affected by other quantities, the corresponding elements of χ' and χ will coincide. In the extreme case where all the geological quantities are independent, X simply becomes the identity matrix.

This rotation provides a basis for muon tomography, as opposed to simple radiography. When combinations of geological quantities χ_i' have been measured, the X matrix can be used to transform them back into individual geological quantities χ_n. Therefore, as happens in (for example) an X-ray CT scanner, quantities that cannot be measured independently can be separated at the data analysis stage.

2.3 Analysis using hypothesis testing

Given orthogonal combinations of geological quantities, it is possible to construct a set of geological hypotheses, each of which describes a possible state of the geology. The complete set of hypotheses $\mathbb{H}$ is assumed to be the Cartesian product

$$\mathbb{H} = \mathbb{H}^1 \times \mathbb{H}^2 \times \cdots \times \mathbb{H}^N, \tag{27}$$

where each $\mathbb{H}^i$ is a set corresponding to the ith combination of geological quantities. (In other words, the elements of $\mathbb{H}$ correspond to every possible combination of the elements of $\mathbb{H}^i$.) Each $\mathbb{H}^i$ has H_i elements, corresponding to H_i possible values of χ'_i. These are indexed as $\mathbb{H}^i_{h_i}$, where the upper index corresponds to a particular combination of geological quantities, and the lower index corresponds to a hypothesized value.

In order to distinguish between these combinations of geological hypotheses, a test statistic for the ith combination is defined as

$$\Psi_i = \sum_m^M \gamma_{im} k_m, \tag{28}$$

where k_m is the set of muon count measurements, and where γ_{im} are weighting coefficients, the optimal values of which are to be determined. The assumption that the test statistic is a linear combination is a loss of generality, but as shown in Appendix C, this form is unlikely to differ greatly from an optimal statistic. For a single combination of geological hypotheses $\mathbb{H}^i_{h_i}$, the test statistic will have an expectation value

$$\psi_{ih_i} = \sum_m^M \gamma_{im} \lambda_m. \tag{29}$$

This form is problematic, as the index h_i corresponds to hypotheses for a single combination of geological quantities, whereas λ_m is dependent on all such combinations. In other words, it is necessary for the left-hand side of the equation to be a function of just $\mathbb{H}^i$, whereas the right-hand side of the equation is in general a function of all $\mathbb{H}$. The problem can be made more explicit by substituting eqs (2), (3) and (21) into eq. (29), to give

$$\psi_{ih_i} = t \sum_m^M \gamma_{im} \tilde{\mu}_m \left(1 + \sum_j^N \zeta'_{jm} \chi'_j \left(\mathbb{H}^j_{h_j}\right)\right), \tag{30}$$

where $\mathbb{H}$ has been split into its components, as per eq. (27). For the test statistic to be useful in isolating the ith combination of geological quantities, all terms in this summation other than that for $i = j$ must equal zero. This means that

$$\sum_m^M \gamma_{im} \zeta'_{jm} = \xi_i \delta_{ij}. \tag{31}$$

It will become apparent that the case $\gamma_{im} = \zeta'_{im}$ is of particular importance, as this corresponds to optimal resolution. (This is stated in eq. (50), and justified in Appendix E.) In this case, eq. (31) becomes eq. (23), which is satisfied by the orthogonalization described in Section 2.2. Therefore eq. (30) can be reduced to

$$\psi_{ih_i} = t \sum_m^M \gamma_{im} \tilde{\mu}_m \left(1 + \zeta'_{im} \chi'_i \left(\mathbb{H}^i_{h_i}\right)\right) \tag{32}$$

which, as required, depends on only the ith combination of geological quantities. The test statistic can be analysed by assuming the measurement to imply $\mathbb{H}^i_{h_i}$, so long as

$$b_{i,h_i-1} < \Psi_i \leq b_{ih_i}, \tag{33}$$

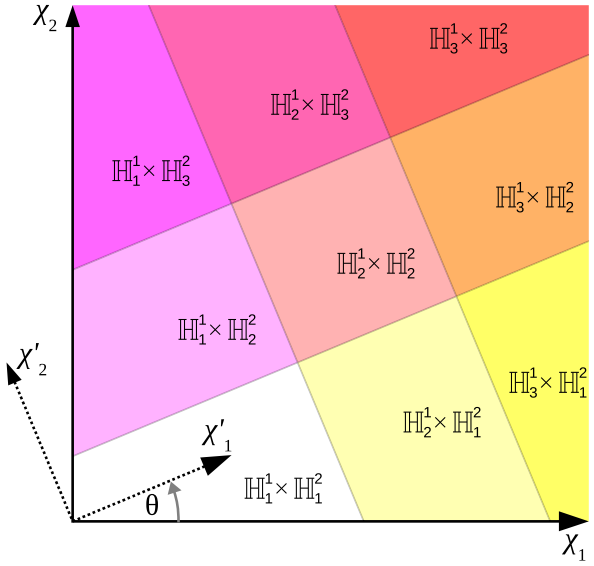

Figure 3. Schematic of hypothesis testing in an orthogonal basis. A pair of geological quantities χ_1 and χ_2 have been transformed into orthogonal combinations χ'_1 and χ'_2, in what amounts to rotation by angle θ. Three different hypotheses for both χ'_1 and χ'_2 give nine hypotheses overall. These map onto hypotheses in χ_1 and χ_2 via the rotation.

where b_{ih_i} are threshold values. A schematic of hypothesis testing is given in Fig. 3.

The probability of the correct hypothesis being chosen for the ith combination of geological quantities is

$$Q_i = \sum_{h_i=1}^{H_i} \mathrm{P}\left(b_{i,h_i-1} < \Psi_i \leq b_{ih_i} | \mathbb{H}^i_{h_i}\right) \mathrm{P}\left(\mathbb{H}^i_{h_i}\right), \tag{34}$$

where $\mathrm{P}\left(\mathbb{H}^i_{h_i}\right)$ is the prior probability of the geological hypothesis $\mathbb{H}^i_{h_i}$. It may be possible to extend this use of prior probabilities into a Bayesian inference scheme, but doing so is beyond the scope of this paper.

It is assumed that the set of hypotheses is collectively exhaustive, so that

$$\mathrm{P}\left(\Psi_i \leq b_{i0} | \mathbb{H}^i_1\right) = 0 \tag{35}$$

$$\mathrm{P}\left(\Psi_i \leq b_{iH_i} | \mathbb{H}^i_{H_i}\right) = 1. \tag{36}$$

Therefore, eq. (34) can be rewritten as

$$Q_i = \mathrm{P}\left(\mathbb{H}^i_{H_i}\right) + \sum_{h_i=1}^{H_i-1} \left[\mathrm{P}\left(\Psi_i \leq b_{ih_i} | \mathbb{H}^i_{h_i}\right) \mathrm{P}\left(\mathbb{H}^i_{h_i}\right) - \mathrm{P}\left(\Psi_i \leq b_{ih_i} | \mathbb{H}^i_{h_i+1}\right) \mathrm{P}\left(\mathbb{H}^i_{h_i+1}\right)\right]. \tag{37}$$

When large muon counts are expected, it follows that

$$\sum_m^M \lambda_m \gg 1. \tag{38}$$

Consequently, as shown in Appendix D, the probability mass function of Ψ_i can be approximated by the normal probability density function, even if the individual distributions of k_m cannot. This will have mean ψ_{ih_i} and standard deviation σ_{ih_i}, with the latter given by

$$\sigma^2_{ih_i} = \sum_m^M \gamma^2_{im} \lambda_m. \tag{39}$$

The cumulative distribution function for a random variable Ψ, normally distributed with expectation ψ and standard deviation σ, is given by

$$\mathrm{P}\left(\Psi \leq c\right) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{c-\psi}{\sqrt{2\sigma^2}}\right)\right], \tag{40}$$

where the error function is defined here as

$$\mathrm{erf}\left(x\right) \equiv \frac{2}{\sqrt{\pi}}\int_0^x \exp\left(-y^2\right)\mathrm{d}y. \tag{41}$$

Therefore, making the normal approximation allows eq. (37) to be written as

$$\begin{aligned} Q_i &= \mathrm{P}\left(\mathbb{H}^i_{H_i}\right) \\ &+ \frac{1}{2}\sum_{h_i=1}^{H_i-1}\left[\mathrm{erf}\left(\frac{\beta_{ih_i} - \sum_m^M \gamma_{im}\tilde{\mu}_m\left(1+\zeta'_{im}\chi'_i\left(\mathbb{H}^i_{h_i}\right)\right)}{\sqrt{2\sum_m^M \gamma_{im}^2\tilde{\mu}_m\left(1+\zeta'_{im}\chi'_i\left(\mathbb{H}^i_{h_i}\right)\right)}}\sqrt{t}\right)\mathrm{P}\left(\mathbb{H}^i_{h_i}\right)\right. \\ &\left. - \mathrm{erf}\left(\frac{\beta_{ih_i} - \sum_m^M \gamma_{im}\tilde{\mu}_m\left(1+\zeta'_{im}\chi'_i\left(\mathbb{H}^i_{h_i+1}\right)\right)}{\sqrt{2\sum_m^M \gamma_{im}^2\tilde{\mu}_m\left(1+\zeta'_{im}\chi'_i\left(\mathbb{H}^i_{h_i+1}\right)\right)}}\sqrt{t}\right)\mathrm{P}\left(\mathbb{H}^i_{h_i+1}\right)\right], \end{aligned} \tag{42}$$

where the division thresholds have been written in a time independent form β_{ih_i}, defined by

$$b_{ih_i} = t\beta_{ih_i}. \tag{43}$$

The numerators within the error functions of eq. (42) contain a subtraction which is critically dependent on the value of $\eta_m\left(\mathbb{H}^i_{h_i}\right)$. Conversely, the denominators are such that the approximation

$$1 + \zeta'_{im}\chi'_i \approx 1 \tag{44}$$

can safely be made. [A similar approximation was made in the derivation of eq. (19).] Using this approximation, and choosing β_{ih_i} to be the mean of the surrounding expectation values, such that

$$\beta_{ih_i} = \sum_m^M \gamma_{im}\tilde{\mu}_m\left(1+\zeta'_{im}\left(\frac{1}{2}\chi'_i\left(\mathbb{H}^i_{h_i+1}\right) + \frac{1}{2}\chi'_i\left(\mathbb{H}^i_{h_i}\right)\right)\right) \tag{45}$$

allows eq. (42) to be rewritten as

$$\begin{aligned} Q_i &= \mathrm{P}\left(\mathbb{H}^i_{H_i}\right) \\ &+ \sum_{h_i=1}^{H_i-1}\mathrm{erf}\left(\frac{\sum_m^M \gamma_{im}\tilde{\mu}_m\zeta'_{im}\left(\chi'_i\left(\mathbb{H}^i_{h_i+1}\right) - \chi'_i\left(\mathbb{H}^i_{h_i}\right)\right)}{\sqrt{8\sum_m^M \gamma_{im}^2\tilde{\mu}_m}}\sqrt{t}\right) \\ &\times \frac{\mathrm{P}\left(\mathbb{H}^i_{h_i}\right) + \mathrm{P}\left(\mathbb{H}^i_{h_i+1}\right)}{2}. \end{aligned} \tag{46}$$

In the special case where $H_i = 2$, or more generally, where $\chi'_i\left(\mathbb{H}^i_{h_i}\right)$ are evenly spaced between the hypotheses, such that

$$z_{im} = \zeta'_{im}\left(\chi'_i\left(\mathbb{H}^i_{h_i+1}\right) - \chi'_i\left(\mathbb{H}^i_{h_i}\right)\right), \tag{47}$$

where z_{im} is the spacing, eq. (46) reduces to

$$\begin{aligned} Q_i &= \mathrm{P}\left(\mathbb{H}^i_{H_i}\right) \\ &+ \mathrm{erf}\left(\frac{\sum_m^M \gamma_{im}\tilde{\mu}_m z_{im}}{\sqrt{8\sum_m^M \gamma_{im}^2\tilde{\mu}_m}}\sqrt{t}\right)\left(1 - \frac{\mathrm{P}\left(\mathbb{H}^i_1\right) + \mathrm{P}\left(\mathbb{H}^i_{H_i}\right)}{2}\right) \end{aligned} \tag{48}$$

At this point, the coefficients γ_{im} can be chosen so that the probability of obtaining the correct answer Q_i is maximized. As the error function is monotonically increasing, this maximization is equivalent to the maximization of

$$q_i = \frac{\left(\sum_m^M \tilde{\mu}_m z_{im}\gamma_{im}\right)^2}{\sum_m^M \tilde{\mu}_m\gamma_{im}^2} \tag{49}$$

over with respect to γ_{im}, for a given index i.

It is shown in Appendix E that this occurs when $\gamma_{im} \propto z_{im}$. A proportional, rather than an absolute solution is to be expected, as the converse would imply that it is possible to extract differing amounts of information from a set of numbers by multiplying them all by a common constant. Due to the unimportance of the coefficient of proportionality, it can be arbitrarily chosen to be the spacing between hypotheses. From eq. (47), this gives

$$\gamma_{im} = \zeta'_{im} \tag{50}$$

thus justifying the assumption made when deriving eq. (32). Substituting eq. (50) into eq. (28) gives the optimum test statistic as

$$\Psi_i = \sum_m^M \zeta'_{im}k_m. \tag{51}$$

The optimized value of Q_i can then be obtained by substituting eqs (47) and (50) into eq. (48), giving

$$\begin{aligned} Q_i &= \mathrm{P}\left(\mathbb{H}^i_{H_i}\right) \\ &+ \mathrm{erf}\left(\sqrt{\frac{1}{8}\sum_m^M \tilde{\mu}_m\zeta'^2_{im}\left(\chi'_i\left(\mathbb{H}^i_{h_i+1}\right) - \chi'_i\left(\mathbb{H}^i_{h_i}\right)\right)^2}\sqrt{t}\right) \\ &\times \left(1 - \frac{\mathrm{P}\left(\mathbb{H}^i_1\right) + \mathrm{P}\left(\mathbb{H}^i_{H_i}\right)}{2}\right). \end{aligned} \tag{52}$$

Therefore, given some acceptable value for Q_i, the time taken to distinguish between hypotheses is

$$t_i = \frac{\tau_i}{\sum_m^M \tilde{\mu}_m\zeta'^2_{im}\left(\chi'_i\left(\mathbb{H}^i_{h_i+1}\right) - \chi'_i\left(\mathbb{H}^i_{h_i}\right)\right)^2}, \tag{53}$$

where τ_i is a constant of proportionality.

The right-hand sides of eqs (19) and (53) are effectively reciprocals of one another. This suggests that the two approaches are equivalent, as the former is data per unit time, whereas the latter is time required to acquire a particular piece of data. The relation can be made more explicit by combining eqs (4) and (21) to give

$$\sum_n^N \zeta_{nm}\chi_n = \sum_i^N \zeta'_{im}\chi'_i \tag{54}$$

which allows eq. (19) to be written as

$$C_\mathrm{b} = \frac{\ln\left(2\right)}{8}\sum_m^M \tilde{\mu}_m\left(\sum_i^N \left|\zeta'_{im}\right|\Delta\chi'_i\right)^2. \tag{55}$$

The principal difference is that eq. (55) considers all hypotheses over the maximum possible range, and so provides the total data rate, whereas eq. (53) applies to a particular combination of hypotheses, over some subrange, in which case only the data contributing to the resolution of those hypotheses is included.

The prefactor in eq. (55) can be used to set τ_i in eq. (53), giving

$$t_i = \frac{1}{\frac{\ln(2)}{8}\sum_m^M \tilde{\mu}_m\zeta'^2_{im}\left(\chi'_i\left(\mathbb{H}^i_{h_i+1}\right) - \chi'_i\left(\mathbb{H}^i_{h_i}\right)\right)^2} \tag{56}$$

as the mean time required to acquire the data bit to separate hypotheses. When there are two hypotheses, this can be written more conveniently as

$$t_i = \frac{1}{\frac{\ln(2)}{8}\sum_m^M \tilde{\mu}_m\zeta'^2_{im}\left(\chi'^\mathrm{A}_i - \chi'^\mathrm{B}_i\right)^2}, \tag{57}$$

where χ'^A_i and χ'^B_i represent the values of χ'_i for hypotheses labelled A and B, respectively.

2.4 Sorting observations by direction of arrival

So far, the muon counts have been treated in a very general sense, with little consideration as to which count a given muon observation is assigned to. If the muon observations are sorted by direction of arrival, where each set of directions covers a solid angle element $\Delta\Omega_m$, then μ_m can be split up as

$$\mu_{\mathrm{m}} = J_m a_m \Delta\Omega_m, \tag{58}$$

where J_m is the muon flux density, and a_m is the effective detector area. Likewise, the expected muon counts can be written as

$$\lambda_m = J_m a_m \Delta\Omega_m t. \tag{59}$$

Substituting eq. (58) into eq. (19) gives

$$C_{\mathrm{b}} = \frac{\ln(2)}{8} \sum_m^M \left(\sum_n^N |\zeta_{nm}| \, \Delta\chi_n \right)^2 \tilde{J}_m a_m \Delta\Omega_m, \tag{60}$$

where $\tilde{J}_m$ corresponds to J_m for the $\mu_{\mathrm{m}} = \tilde{\mu}_{\mathrm{m}}$ case. Similarly, eq. (57) becomes

$$t_i = \frac{1}{\frac{\ln(2)}{8} \sum_m^M \zeta_{im}'^2 \left(\chi_i'^{\mathrm{A}} - \chi_i'^{\mathrm{B}} \right)^2 \tilde{J}_m a_m \Delta\Omega_m} \tag{61}$$

and the ζ_{nm} coefficients can be derived as

$$\zeta_{nm} = \left[\frac{1}{J_m} \frac{\partial J_m}{\partial \chi_n} \right]_{\chi_{1\cdots N}=0}. \tag{62}$$

The summations in eqs (60) and (61) can be thought of as being approximations to integrals with respect of Ω. If $\Delta\Omega_m$ becomes infinitesimal, then these equations will indeed contain a continuously varying $\tilde{J}$ integrated over all of the solid angle.

As an aside, when constructing an image of muon flux density with direction of arrival, sufficient information must exist to distinguish between regions of the image. This limits the meaningful angular resolution, regardless of the accuracy of the detectors. In Appendix F, this limiting resolution is estimated to be of the order

$$\phi_m^{\mathrm{lim}} \approx \frac{1}{\Delta\eta_m \sqrt{J_m a_m t}}, \tag{63}$$

where $\Delta\eta_m$ is the maximum likely range of η_m. Due to the very crude approximations made, this should not be interpreted as a division between necessary and excessive resolution. It may, however, provide a useful guideline for detector specifications.

3 MODELLING MUON FLUX DENSITY

In order to apply the data analysis technique described above, it is necessary to have a model of the geology, to calculate the muon flux densities J_m for different geological scenarios. While there are a range of muon simulation software packages, including Geant (Agostinelli 2003), MMC (Chirkin & Rhode 2004), MUM (Sokalski *et al.* 2001) and MUSIC (Antonioli *et al.* 1997; Kudryavtsev *et al.* 1999; Kudryavtsev 2009), they work on a muon-by-muon basis, and so can't provide the values of J_m directly. Furthermore, the need to differentiate these values with respect to the geological quantities presents a problem, in that numerical differentiation is very noise sensitive. It is therefore useful to perform modelling based on the expectation values themselves.

3.1 Calculation of survival probability function

As a preliminary to calculating muon flux densities, a realistic survival probability function, $\Xi(s)$, is required. This is a function of propagation distance s, and gives the probability that a randomly selected muon will reach that distance. The MUSIC code was used to simulate the passage of muons through rock, taking into account ionization, bremsstrahlung, electron-positron pair production and inelastic scattering between muons and atomic nuclei.

The propagation of an individual muon is dependent on its starting energy. Therefore, the survival probability function will depend on the distribution of muon energies at ground level. To take this into account, the Gaisser parametrization (Gaisser 1990) was used. This is a semi-empirical model of high energy muon flux at ground level, as a function of muon energy E and zenith angle θ. This gives the mean arrival rate of $\geq$100 GeV muons per unit area, per unit solid angle, per unit energy as

$$g(E,\theta) = \begin{cases} 0 & ; \quad E < 100 \text{ GeV} \\ \frac{1400}{\mathrm{m}^2 \text{ s sr GeV}} \left(\frac{E}{\mathrm{GeV}} \right)^{-2.7} \left(F_\pi(E,\theta) + F_K(E,\theta) \right); & E \geq 100 \text{ GeV} \end{cases}, \tag{64}$$

where the fact that <100 GeV muons are not being considered is made explicit. This is done due to both the limit of applicability of the Gaisser parametrization, and the low probability of lower energy muons penetrating substantially into the rock. In the subsequent analysis, muon starting at lower energies don't enter the statistics.

The $E^{-2.7}$ power law corresponds to the energy spectrum of the cosmic ray nucleons entering the earth's atmosphere, and the functions F_π and F_K correspond to the two principal muon producing reactions, which involve the intermediate production of a pion or a kaon. They are given by

$$F_\pi(E,\theta) = \frac{1}{1 + \frac{1.1 E \cos\theta}{115 \text{ GeV}}} \tag{65}$$

$$F_K(E,\theta) = \frac{0.054}{1 + \frac{1.1 E \cos\theta}{850 \text{ GeV}}} \tag{66}$$

To demonstrate the method, survival probability functions were calculated for 'standard rock', a rock model consisting of atoms with proton number 11 and mass number 22, which are approximate averages for the elements present in common rock forming minerals (Groom *et al.* 2001). Standard rock has a density of 2.65 g cm^{-3}, which again is a representative value. MUSIC supports arbitrary chemical composition (as do other simulation packages), and so any rock type can in principle be simulated.

The reason for using standard rock is that for the geology being considered, no accurate data for elemental composition exists. While the composition is known on a mineralogical basis, the exact mineral ratios are not known. Furthermore, many of the minerals exist as series, whereby substances with different elemental compositions are assigned the same mineral name. This problem also affects conventional Monte Carlo methods, and so should not be considered a limitation of this particular method.

As only a one time calculation was required, 1 million simulations were performed, thus minimizing random error. In order to make efficient use of these simulations, the starting energies were chosen using the method described in Appendix G. The calculated values of $\Xi(s)$, for a selection of zenith angles are shown in Fig. 4.

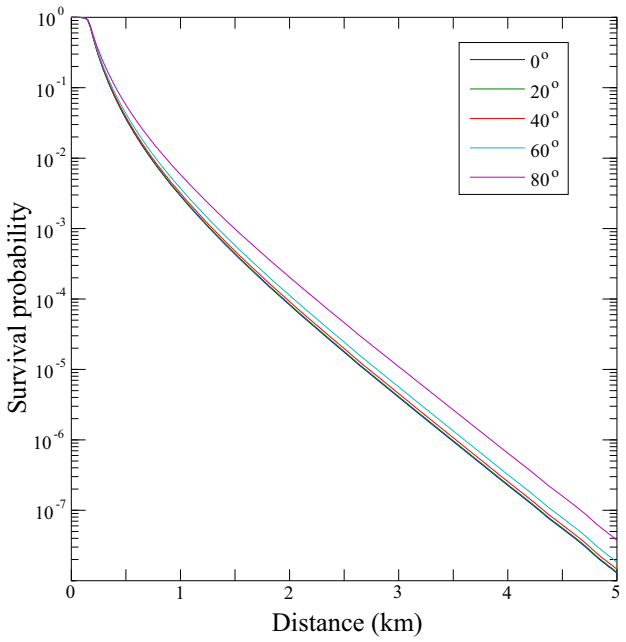

Figure 4. The probability that a randomly chosen ≥100 GeV cosmic ray muon will penetrate a given distance of standard rock, assuming the Gaisser model of ground level flux, shown for a selection of zenith angles. The flat region preceding 200 m is a consequence of the 100 GeV energy cut-off. Further than this, the probability of any <100 GeV muon surviving is negligible, and the probability function is quasi-exponential.

3.2 Probability ray tracing

The muon flux density below ground can be calculated by performing ray tracing based on the above survival functions. The ray tracing makes the approximation that the muons travel in a straight line, and that deflections, both in angular direction and in position perpendicular to the original ray path can be neglected. This is an acceptable approximation for the high energy muons being considered here (Lesparre *et al.* 2010), but becomes problematic at lower energies.

The material in each cell is represented by its bulk density ρ. This is an approximation, as in reality the survival probability function depends somewhat on elemental composition. A further approximation is that when the density of the rock differs from that of some reference rock (which in the demonstration below is standard rock) the difference in muon survival is largely due to the increase or reduction of the total material encountered on route. Again, this is a reasonable approximation for high-energy muons, as the number of particle interactions is more important than the rate at which they occur. (This would not be true for low energy muons, as the effect of muon decay via the weak nuclear force would be significant. However, for the very high energy muons being considered, relativistic time dilation reduces the observed rate of their decay by a factor of 1000 or more.)

To implement this approximation, the muon survival probability function is represented as a function of effective distance $\tilde{s}$. This is defined as

$$\tilde{s} = \int_{s_{\rm in}}^{s_{\rm out}} \frac{\rho\,(s)}{\rho_0}\,{\rm d}s, \tag{67}$$

where $\rho_0 = 2.65$ g cm^{-3} is the density of standard rock. By definition, muons encountering the same integrated density of material will have traveled the same effective distance. When all the material encountered has the same density as the reference rock, the real and effective distances will be the same.

In the mth direction, the muon arrival rate per unit area, per unit solid angle, per unit energy, j_m, is given by multiplying the corresponding rate at ground level by the survival probability function. This gives

$$j_m = g\,(E, \theta_m)\ \ \Xi\,(\tilde{s}_m, \theta_m)\,, \tag{68}$$

where θ_m and $\tilde{s}_m$ are the zenith angle and effective distance for that particular direction. Integrating these spectra with respect to energy gives the muon arrival rate per unit area per unit solid angle at the detector as

$$J_m = \int_0^\infty j_m\ {\rm d}E \tag{69}$$

and the muon arrival rate per unit area per unit solid angle at ground level as

$$G\,(\theta_m) = \int_0^\infty g\,(E, \theta_m)\,{\rm d}E. \tag{70}$$

As per Section 3.1, these functions only apply to muons with an energy of ≥100 GeV at sea level. This becomes unimportant beyond a depth of about 200 m, as the survival probability of lower muons energy becomes negligible.

Integrating eq. (68) over all energies gives

$$J_m = G\,(\theta_m)\ \ \Xi\,(\tilde{s}_m, \theta_m)\,. \tag{71}$$

Ray traced images, such as those shown in Section 4 can then be generated by evaluating eq. (71) in all directions. The implementation of this ray tracing is described in Appendix H.

4 APPLICATION TO GEOLOGICAL STORAGE OF CARBON DIOXIDE

To demonstrate the value of the methods derived in this paper, they have been applied to a model of a geological CO_2 storage scenario. This is based around real geophysical data for the Boulby Mine, in Yorkshire, England, which extends laterally for 7 km beneath the sea at depths of between about 750 m and 1.4 km. While CO_2 storage is not planned for this site, it is chosen because a deep science facility [the Boulby Underground Laboratory (Murphy & Paling 2012)] is already located there, and the geology is very similar to that known to exist at several of the preferred storage sites in the North Sea (Hedley *et al.* 2013).

This demonstration is a useful way of testing the equations derived in the previous sections with realistic numbers, rather than just abstract values. It concerns radiography rather than full tomography. However, the numerical values will be of a similar order of magnitude, and so assessing the method as applied to radiography will provide an idea as to its usefulness when applied to tomography.

4.1 Model of geology and carbon dioxide injection

The geology around the Boulby mine consists of five principal layers. The first layer, starting from the sea bed is Liassic shale, the second layer is Triassic Mercia mudstone and the third layer is Triassic Bunter sandstone. The Boulby mine workings straddle the boundary between a fourth and a fifth layer, the former of which consists of Upper-Permian sylvite and polyhalite and the latter of which consists of Upper-Permian halite, sylvite, polyhalite and other evaporite minerals. The model includes the sea itself as a

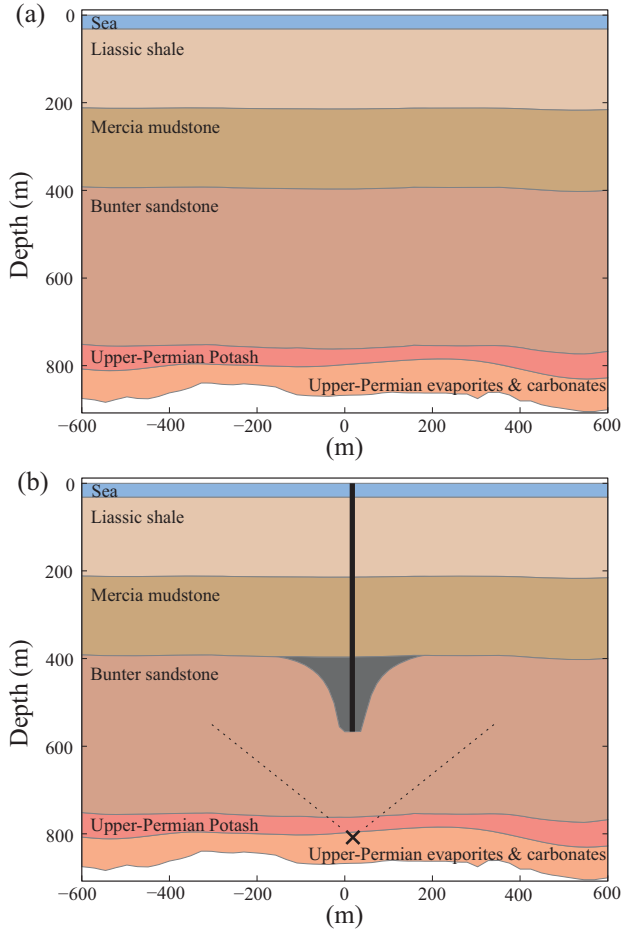

Figure 5. Cross section of the Boulby mine geology. The mine workings straddle the boundary between the two upper-Permian layers. (a) At present. (b) After a (hypothetical) injection of supercritical CO_2 into the Bunter sandstone, creating a plume capped by the Mercia mudstone. The vertical line denotes the well, and the black cross gives the location of a possible muon detector (with the dashed lines showing the field of view used for ray tracing).

zeroth layer, having a depth of 32 m. A cross section of this geology is shown in Fig. 5(a).

The carbon capture scheme is assumed to operate by injecting supercritical CO_2 fluid into the Bunter sandstone. This is porous and permeable, allowing CO_2 to move from the injection site, by displacing brine from the rock's pore system. The Mercia mudstone above the Bunter sandstone has a much lower permeability, and should therefore act as a seal, preventing CO_2 from leaking upwards towards the sea. The purpose of the muon tomography is to determine whether or not the expected processes are happening in practice.

The initial movement of the CO_2 and brine was modelled using a method based on (Mathias *et al.* 2009). (The later sequestration stages involving CO_2 dissolution and mineral trapping mechanisms, where chemical reactions take place, are not considered in this demonstration.) The Bunter layer is now considered as two subregions, this division being equivalent to the difference between Parts A and B of Fig. 5. The upper subregion is assumed to be filled to the maximum extent with CO_2 (i.e. the pore space is filled with CO_2 plus a specified fraction of immobile brine), and the lower subregion is assumed to be unaffected by the injection (i.e. all pore space contains brine). The density of these regions, and the shape

Table 1. Physical parameters used in the CO_2 flow model. These correspond to the Boulby site.

Parameter	Symbol	Value	Units
Brine density	ρ_w	1100	$kg\,m^{-3}$
Brine viscosity	μ_w	900	$\mu Pa\,s$
CO_2 density	ρ_n	720	$kg\,m^{-3}$
CO_2 viscosity	μ_n	60	$\mu Pa\,s$
Rock density	ρ_s	2670	$kg\,m^{-3}$
Rock porosity	ϕ	0.15	–
CO_2 relative permeability	k_{rn}	0.3948	–
Brine relative permeability	k_{rw}	1	–
Brine residual saturation	S_{rw}	0.4438	–
Well/reservoir height	H	170	m
CO_2 mass injection rate	M_n	20	$kg\,s^{-1}$

of the surface between them are calculated from a simple model of fluid flow through porous media, as detailed in Appendix I. This includes numerous simplifications (as described in the appendix) but it provides a reasonable example of the general scenario that the method will be applied to. More sophisticated models have been applied in other papers, but at the expense of far higher computational overheads. (Klinger *et al.* 2015; Gluyas *et al.* 2018).

The calculations were performed using real values for the Boulby site, as displayed in Table 1. The rock density, porosity, and permeability characteristics are measured values for the sandstone. The fluid properties were derived using equations of state (Rowe & Chou 1970; Bell at al. 2014). A CO_2 injection rate of 20 $kg\,s^{-1}$ was chosen, which is unexceptional for the range (3–120 $kg\,s^{-1}$) considered practical (Mathias *et al.* 2011a). The well is assumed to protrude 170 m into the Bunter sandstone, reaching a total depth below sea level of 569 m. Fig. 5(b) shows a cross section of the geology after 1 yr of CO_2 injection.

4.2 Results

Fig. 6 shows muon flux densities for the cases of no CO_2 injection, and 1 yr of CO_2 injection, respectively. The detector is assumed to be located within the mine workings directly beneath the well at a depth below sea level of 808 m. The difference between the images is slight, as the muon attenuation is dominated by the unchanged geological factors. Fig. 7(a) shows the relative change in muon flux density. This change is always positive, as the supercritical CO_2 is less dense than the brine it replaces, reducing the effective propagation length of the muons.

If the difference between the cases is parametrized by the extent to which the CO_2 has been retained, with $\chi_n = 0$ being no retention (i.e. the no injection case), and $\chi_n = 1$ being full retention (i.e. the expected 1 yr case), then Figs 6(a) and 7(a) are, respectively, plots of $\tilde{J}_m$ and ζ_{nm} (the latter being for the single parameter $n = 1$). If the object of the data analysis is to distinguish between these cases, such that $\Delta\chi_n = 1$, then the total information rate can be calculated via eq. (60). The summand of this Equation gives the information rate from each direction, as is shown in Fig. 7(b).

4.3 Comparison of test statistics

This modelling can be used to compare the new test statistic Ψ_i, as given by

$$\Psi_i = \sum_m^M \zeta'_{im} k_m \tag{72}$$

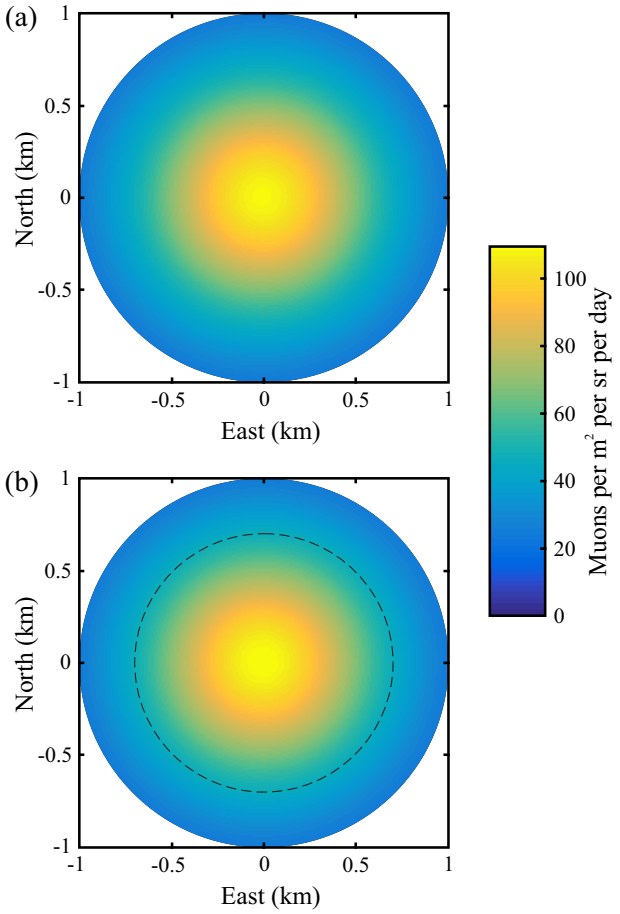

Figure 6. Calculated muon flux density at the location of the detector. Displayed using Gnomonic projection onto a plane at surface level, where the axes give the horizontal distance from the detector. Shown (a) before CO_2 injection and (b) after 1 yr of CO_2 injection. The dashed line shows the outer extent of the CO_2 plume.

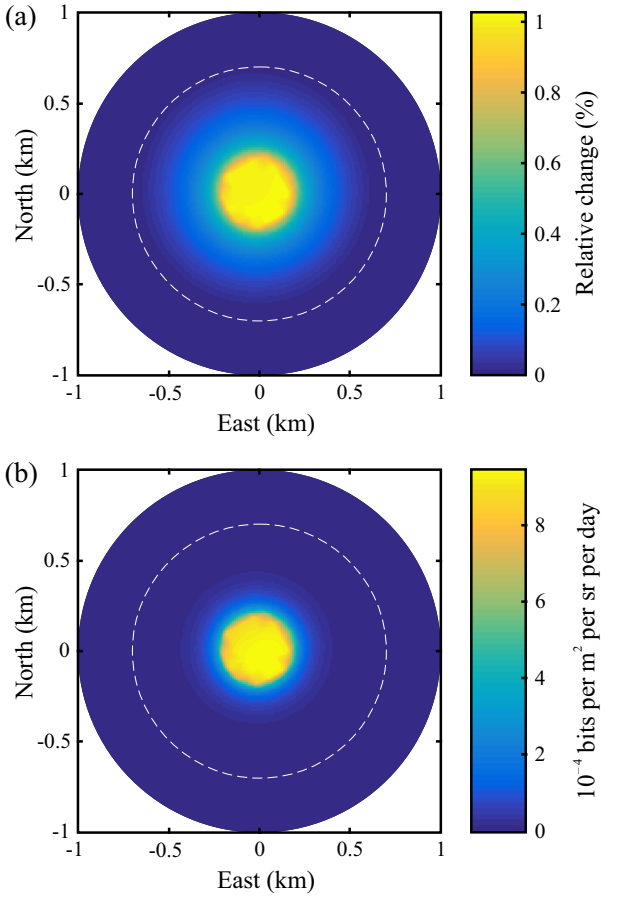

Figure 7. Quantities derived from the data in Fig. 6. (a) Relative change in muon flux density. (b) Theoretical information acquisition rate, per unit detector area per unit solid angle.

with the more obvious test statistic, that is the sum of muons arriving from trajectories passing through some object of interest, as given by

$$\Phi_i = \sum_m \delta_m k_m, \tag{73}$$

where δ_m is assigned a value of 1 if the mth ray path passes through the object, and is 0 otherwise. In the context of Figs 6 and 7, the value of δ_m is 1 within the dashed line and 0 outside.

The counts k_m have expectation value

$$\lambda_m = J_m a_m \Delta\Omega_m t, \tag{74}$$

where J_m is derived from the above modelling, $a_m \Delta\Omega_m$ depends on the detector and t is observation time. Therefore, the expectation values for Ψ_i and Φ_i (denoted by ψ_i and ϕ_i, respectively) are given by

$$\psi_i = \sum_m \zeta'_{im} J_m a_m \Delta\Omega_m t \tag{75}$$

$$\phi_i = \sum_m \delta_m J_m a_m \Delta\Omega_m t. \tag{76}$$

The same for typical values of muon flux density are also required. When monitoring CO_2 storage, these typical values are most usefully chosen as those corresponding to the geology prior to CO_2 injection. The expectation values for the prior to injection case are therefore given by

$$\tilde{\psi}_i = \sum_m \zeta'_{im} \tilde{J}_m a_m \Delta\Omega_m t \tag{77}$$

$$\tilde{\phi}_i = \sum_m \delta_m \tilde{J}_m a_m \Delta\Omega_m t, \tag{78}$$

where $\tilde{J}_m$ are likewise the values of J_m prior to injection. In the limit $\sum_m \lambda_m \gg 1$, the quantities Ψ_i and Φ_i are both normally distributed, where the null hypothesis has standard deviations given by

$$\sigma^2_{\tilde{\psi}_i} = \sum_m^M \zeta'^2_{im} \tilde{J}_m a_m \Delta\Omega_m t \tag{79}$$

$$\sigma^2_{\tilde{\phi}_i} = \tilde{\phi}_i. \tag{80}$$

In the case of Φ_i, this is a basic property of the Poisson distribution. The equivalent result for Ψ_i follows from Appendix D.

As the distribution for the null hypothesis is normal, with known mean and standard deviation, the relevant statistical test is the Z-test. As the muon rate is expected to increase with CO_2 injection, an upper-tailed Z-test should be used. In actual fact, analysis of real data probably would not proceed like this, as instead of using values of $\tilde{J}_m$ from modelling, the null hypothesis expectation values will be estimated from muon observations taken during a learning period prior to CO_2 injection. Instead, the modelling will only enter the analysis via the ζ'_{im} coefficients. The reasons for this difference are discussed in Section 5.

That said, this hypothetical Z-test remains a useful way of evaluating the information gathering capabilities of the revised test statistic when realistic numbers are used. For a test statistic x, the Z-score is calculated as

$$z = \frac{x - \mu}{\sigma}, \tag{81}$$

where μ and σ are the mean and standard deviation of the null hypothesis. If the null hypothesis is to be rejected for a given p-value, then z must exceed a certain threshold, which is dependent on p. If, for example, a p-value of less than 0.05 is required, then the threshold value of z is 1.645. The two test statistics can be converted into a Z-score as

$$z_\Psi = \frac{\Psi_i - \tilde{\psi}_i}{\sigma_{\psi_i}} \tag{82}$$

$$z_\Phi = \frac{\Phi_i - \tilde{\phi}_i}{\sigma_{\phi_i}}. \tag{83}$$

The effectiveness of these test statistics in detecting an alternative hypotheses can be assessed by calculating their expectation values in the case that the hypothesis is true. These are given by

$$\bar{z}_\Psi = \frac{\psi_i - \tilde{\psi}_i}{\sigma_{\psi_i}} \tag{84}$$

$$\bar{z}_\Phi = \frac{\phi_i - \tilde{\phi}_i}{\sigma_{\phi_i}}. \tag{85}$$

The natural way of comparing these values is to compare their relation to the observation time and the effective area of the muon detectors. If these are chosen differently for different test statistics, then dividing eq. (85) by eq. (84), substituting in the definitions for mean and standard deviations, and then re-arranging gives

$$t_\Psi a_\Psi = \left(\frac{\bar{z}_\Phi}{\bar{z}_\Psi}\right)^2 t_\Phi a_\Phi, \tag{86}$$

where t_Ψ and t_Φ are the observation times, and a_Ψ and a_Φ are scaling factors for the effective detector areas. For a given choice of p-value, this relates the expected requirement for observation times. Making this calculation from the above results gives $t_\Psi a_\Psi \approx 0.48\, t_\Phi a_\Phi$. This means that the required observation time required to achieve a certain result can be approximately halved, or that the detector area required to obtain that result in the same amount of time can be approximately halved. Alternatively, it can be thought of as approximately doubling the effective sensitivity of the detectors.

5 SUMMARY AND CONCLUSIONS

Methods to analyse the information content of low-contrast muon radiography have been developed, as have means to optimally analyse available data. While this problem has been analysed in terms of producing images with a specified contrast (Lesparre *et al.* 2010), the analysis described here considers the absolute upper bounds to data acquisition, and provides means to extract data in these extremes.

In order to make use of the method, it is first necessary to obtain the coefficients ζ_{nm}, which represent how the muon flux density in the mth direction (out of M) depends on the nth geological quantity (out of N) to be determined. Given a set of anticipated mean muon flux densities J_m, these coefficients can be calculated as

$$\zeta_{nm} = \left[\frac{1}{J_m}\frac{\partial J_m}{\partial \chi_n}\right]_{\chi_{1\cdots N}=0}, \tag{87}$$

where χ_n are the geological quantities to be measured. The total data rate obtainable by a muon detector can then be calculated as

$$C_\mathrm{b} = \frac{\ln(2)}{8}\sum_m^M \left(\sum_n^N |\zeta_{nm}|\,\Delta\chi_n\right)^2 \tilde{J}_m a_m \Delta\Omega_m, \tag{88}$$

where $\Delta\chi_n$ are the maximum ranges between feasible values of χ_n, where a_m is the effective detector area in that direction, and where $\Delta\Omega_m$ is a solid angle element. The quantity $\tilde{J}_m$ is some typical value of J_m, which depending on how the method is applied, may correspond to the null hypothesis, or a starting value, or both. When multiple detectors are used, the data rates simply add, and so the quantities J_m and ζ_{nm} can be partitioned into subsets corresponding to the individual detectors.

Notably, any linear scaling of χ_n does not matter, as the consequent scaling of ζ_{nm} will cancel out in eq. (88). Even the dimensions of χ_n do not matter, and so the geological quantities can be any form of variation in any unit, so long as they are approximately linearly related to the muon flux density.

The information rate is linearly related to the flux density J_m, but quadratically related to the contrast terms $\Delta\chi_n$. These relations are useful in detector position optimization, as they show that it is important to view as many muons as possible, but even more important to achieve an arrangement where the difference between the expectation values of the count rates is maximized. A similar result was discovered by Lesparre *et al.* (2010), for the purpose of achieving a given contrast in an image of muon flux density.

In certain cases, the geological quantities to be measured will overlap, for example by a muon passing through multiple regions, when these regions are to be measured independently. In such cases, it is necessary to convert these quantities into an orthogonal basis, such that

$$\chi'_i = \sum_n^N X_{in}\chi_n \tag{89}$$

$$\zeta'_{im} = \sum_n^N X^{-1}_{in}\zeta_{nm}, \tag{90}$$

where X is the rotation matrix described in Section 2.2. This will exist so long as the expected change in muon observations caused by variation in one geological quantity is not identical to that caused by variation in any other geological quantity, or a linear combination thereof.

When the geological quantities are independent, X becomes the identity matrix, and for only a single quantity, it simply becomes one. The X matrix is a necessary foundation for muon tomography, in which multiple detectors are used to infer three dimensional structure. Such tomography, however, has yet to be done using real data.

In order to differentiate between geological hypotheses, it is necessary to calculate a test statistic, which can be compared to modelled values. The most obvious choice of test statistic is the sum of muon counts from all appropriate directions. However, an improved test statistic is

$$\Psi_i = \sum_m^M \zeta'_{im} k_m, \tag{91}$$

where k_m are the counts from the mth direction. This will have an expected value

$$\psi_i = \sum_m^M \zeta'_{im}\lambda_m, \tag{92}$$

where for an observation time t, the expectation values of the counts, λ_m are given by

$$\lambda_m = J_m a_m \Delta\Omega_m t. \tag{93}$$

This form of test statistic is also a total, but the count for each direction has been weighted by the corresponding ζ'_{im} coefficient. This allows for more efficient data analysis, by better taking into account prior information. The time to resolve between two hypotheses, labelled A and B, can be estimated as

$$t_i = \frac{1}{\frac{\ln(2)}{8}\sum_m^M \zeta_{im}'^2 \left(\chi_i'^{\mathrm{A}} - \chi_i'^{\mathrm{B}}\right)^2 \tilde{J}_m a_m \Delta\Omega_m}, \tag{94}$$

where $\chi_i'^{\mathrm{A}}$ and $\chi_i'^{\mathrm{B}}$ are the corresponding values of χ_i for the two hypotheses.

In order to apply this method, a form of probability based muon flux modelling has been developed, which provides J_m directly. In contrast with conventional Monte Carlo based methods, it lacks random error, and so the differentiation required to calculate the ζ_{nm} coefficients [as per eq. (87)] can be done without numerical error. It has the additional advantage of being much faster than Monte Carlo techniques.

The technique does however make several approximations. The muons are assumed to travel in straight lines, whereas in reality deflections will be present. Changes in material are modelled only in terms of deviations in density, as referenced from some material with fixed chemical composition. This density change is accounted for by assuming propagation along an effective distance, having the same integrated density as the default material. In reality, differences to muon survival will depend on more than density. This will include the composition of preceding materials, as different materials will reshape the muon energy spectrum in different ways. The implementation neglects low energy muons, and so to work at depths of less than 200 m, models of low energy ($<$100 GeV) muon flux and scattering would have to be incorporated.

Given the above advantages and disadvantages of the ray tracing method, it is best used to complement existing Monte Carlo methods. One approach would be to use it to calculate ζ_{nm} coefficients, and thus construct a superior test statistic, but otherwise use the higher accuracy of Monte Carlo methods, as has been done in other papers (Klinger *et al.* 2015; Gluyas *et al.* 2018). Given that Monte Carlo methods are extremely computationally expensive (Gluyas *et al.* 2018), the ray tracing method has a speed advantage, and so might also be better in situations where large numbers of simulations are required, such as optimizing detector arrangements. Another potential use, given that simpler models tend to provide more reliable extrapolations, is to verify that the more complex Monte Carlo simulations produce reasonable results when applied to novel situations.

When analysing real geophysical data, models of any sort might not be used directly, but instead used to augment other muon observations. This is essential if systematic errors in the model are comparable to the changes being measured. While modelled values for the changes themselves will not be greatly affected (as systematic errors will largely cancel out) absolute values such as $\tilde{J}_m$ will be directly affected, causing similar systematic errors in the calculated expectation values for the test statistics. In cases where changes in the geology are to be measured, this problem can be avoided by making additional observations prior to the main program of observations. Data from this learning phase can then be used to estimate the expectation value of the test statistic in the case where no change has occurred. Expectation values for other cases can then be estimated by applying model derived changes to this baseline value.

The information theory approach has been applied to a case study based on geological CO_2 storage, where muon tomography is used to monitor the emplacement of the injected CO_2. This suggests that using such methods can reduce the detector cross-sectional area required to achieve a certain result by about half. Given the high costs of manufacturing and deploying detectors, this might result in substantial cost reductions, making the methods described in this paper of prime importance to experimental geophysics.

ACKNOWLEDGEMENTS

This work was supported by the former UK Department of Energy and Climate Change (now part of the Department for Business, Energy and Industrial Strategy), and by Premier Oil plc. Geophysical data was provided by Cleveland Potash Limited. The contribution of M. Coleman was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

REFERENCES

Agostinelli, S., 2003. GEANT4a simulation toolkit, *Nucl. Instrum. Methods Phys. Res., Sect. A,* **506,** 250–303.

Alfaro, R. *et al.*, 2003. A muon detector to be installed at the Pyramid of the Sun, *Revista Mexicana de Fisica Supplement,* **49,** 54–59.

Antonioli, P., Ghettim, C., Korolkova, E.V., Kudryavtsev, V.A. & Sartorelli, G., 1997. A three-dimensional code for muon propagation through the rock: MUSIC, *Astropart. Phys,* **7,** 357–368.

Avati, V. *et al.*, 2003. Cosmic multi-muon events observed in the underground CERN-LEP tunnel with the ALEPH experiment, *Astropart. Phys.,* **19,** 513–523.

Bell, I.H., Wronski, J., Quoilin, S. & Lemort, V., 2014. Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library CoolProp, *Ind. Eng. Chem. Res.,* **53,** 2498–2508.

Burkhard, J., 1970. Search for hidden chambers in the pyramids, *Science,* **167,** 832–839.

Carbone, D., Gibert, D., Marteau, J., Diament, M., Zuccarello, L. & Galichet, E., 2013. An experiment of muon radiography at Mt Etna (Italy), *Geophys. J. Int.,* **196,** 633–643.

Chirkin, D. & Rhode, W., 2004. Propagating leptons through matter with Muon Monte Carlo (MMC), arXiv:hep-ph/0407075v2.

Davis, M.H.A., 1980. Capacity and Cutoff Rate for Poisson-Type Channels, *IEEE Trans. Inf. Theory,* **26,** 710–715.

Gaisser, T.K., 1990. *Cosmic Rays and Particle Physics,* Cambridge Univ. Press.

Gluyas, J.G. *et al.*, 2018. Passive, continuous monitoring of carbon dioxide geostorage using muon tomography, *Philos. Trans. R. Soc., A,* **377,** 2137.

Groom, D.E., Mokhov, N.V. & Striganov, S.I., 2001. Muon stopping power and range tables 10 MeV – 100 TeV, *At. Data Nucl. Data Tables,* **78,** 183–356.

Haas, S.H. & Shapiro, J.H., 2002. Capacity of the multiple-input, multiple-output Poisson channel, in *Proceedings of the 2002 IEEE International Symposium on Information Theory*, 30 June–5 July 2002 Lausanne, Switzerland.

Hedley, B.J., Davies, R.J., Mathias, S.A., Hanstock, D. & Gluyas, J.G., 2013. Uncertainties in CO_2 storage capacity, case study from the North Sea, UK, *Greenhouse Gases: Sci. Technol.,* **3,** 212–230.

Jiang, X., Hassan, W.A.A. & Gluyas, J.G., 2013. Modelling and monitoring of geological carbon storage: a perspective on cross-validation, *Appl. Energy,* **112,** 784–792.

Jourde, K., Gibert, D., Marteau, J., de Bremond dArs, J., Gardien, S., Girerd, C. & Ianigro, J., 2016. Monitoring temporal opacity fluctuations of large structures with muon radiography: a calibration experiment using a water tower, *Sci. Rep.,* **6,** 23054.

Kabanov, Yu.M., 1978. The capacity of a channel of the Poisson type, *Teoriya Veroyatnostei i ee Primeneniya,* **23,** 143–147.
Klinger, J. *et al.*, 2015. Simulation of muon radiography for monitoring CO_2 stored in a geological reservoir, *Int. J. Greenhouse Gas Control,* **42,** 644–654.
Kudryavtsev, V.A., Korolkova, E.V. & Spooner, N.J.C., 1999. Narrow muon bundles from muon pair production in rock, *Phys. Lett. B,* 251–256.
Kudryavtsev, V.A., 2009. Muon simulation codes MUSIC and MUSUN for underground physics, *Comput. Phys. Commun.,* 339–346.
Kudryavtsev, V.A., Spooner, N.C., Gluyas, J.G., Fung, C. & Coleman, M.L., 2011. Monitoring subsurface CO2 emplacement and security of storage using muon tomography, *J. Greenhouse Gases,* **11,** 21–24.
Kudryavtsev, V.A., Spooner, N.J.C., Gluyas, J., Fung, C. & Coleman, M., 2012. Monitoring subsurface CO_2 emplacement and security of storage using muon tomography, *Int. J. Greenhouse Gas Control,* **11,** 21–24.
Lesparre, N., Gilbert, D., Marteau, J., Declais, Y., Carbone, D. & Galichet, E., 2010. Geophysical muon imaging: feasibility and limits, *Geophys. J. Int.,* **183,** 1348–1361.
Lesparre, N., Gibert, D., Marteau, J., Komorowski, J., Nicollin, F. & Coutant, O., 2012. Density muon radiography of La Soufrire of Guadeloupe volcano: comparison with geological, electrical resistivity and gravity data, *Geophys. J. Int.,* **190,** 1008–1019.
Mathias, S.A., Hardisty, P.E., Trudell, M.R. & Zimmerman, R.W., 2009. Approximate solutions for pressure buildup during CO_2 injection in brine aquifers, *Transp. Porous Media,* **79,** 265–284.
Mathias, S.A., Gluyas, J.G., González Martínez de Miguel, G.J. & Hosseini, S.A., 2011a. Role of partial miscibility on pressure buildup due to constant rate injection of CO_2 into closed and open brine aquifers, *Water Resour. Res.,* **47,** 12525.
Mathias, S.A., González Martínez de Miguel, G.J., Thatcher, K.E. & Zimmerman, R.W., 2011b. Pressure buildup during CO_2 injection into a closed brine aquifer, *Transp. Porous Media,* **89,** 383–397.
Mathias, S.A., Gluyas, J.G., Gonzalez, G., Bryant, S. & Wilson, D., 2013a. On relative permeability data uncertainty and CO2 injectivity estimation for brine aquifers, *Int. J. Greenhouse Gas Control,* **12,** 200–212.
Mathias, S.A., Gluyas, J.G., Mackay, E. & Goldthorpe, W., 2013b. A statistical analysis of well production rates from UK oil and gas fields—implications for carbon capture and storage, *Int. J. Greenhouse Gas Control,* **19,** 510–518.
Morishima, K. *et al.*, 2017. Discovery of a big void in Khufu's Pyramid by observation of cosmic-ray muons, *Nature,* **552,** 386–390.
Morris, C.L. *et al.*, 2014. Analysis of muon radiography of the Toshiba nuclear critical assembly reactor, *Appl. Phys. Lett.,* **104,** 024110.
Murphy, S. & Paling, S.M., 2012. The boulby mine underground science facility: the search for dark matter and beyond, *Nucl. Phys. News,* **22,** 19–24.
Nagamine, K., Iwasaki, M., Shimomura, K. & Ishida, K., 1995. Method of probing inner-structure of geophysical substance with the horizontal cosmic-ray muons and possible application to volcanic eruption prediction, *Nucl. Instrum. Methods Phys. Res., Sect. A,* **356,** 585–595.
Nakamura, K. *et al.*, 2010. Review of Particle Physics, *J. Phys. G,* **37,** 075021.
Nordbotten, J.M., Celia, M.A. & Bachu, S., 2005. Injection and storage of CO_2 in deep saline aquifers: analytical solution for CO_2 plume evolution during injection, *Transp. Porous Media,* **58,** 339–360.
Nordbotten, J.M. & Celia, M.A., 2006. Similarity solutions for fluid injection into confined aquifers, *J. Fluid Mech.,* **561,** 307–327.
Perry, J. *et al.*, 2013. Imaging a nuclear reactor using cosmic ray muons, *J. appl. Phys.,* **113,** 18.
Ponting, D.K., 1989. Corner point geometry in reservoir simulation, in *Proceedings of the 1st European Conference on the Mathematics of Oil Recovery,* EAGE, doi:10.3997/2214-4609.201411305.
Rowe, A.M. & Chou, J.C.S., 1970. Pressure-volume-temperature-concentration relation of aqueous NaCl solutions, *J. Chem. Eng. Data,* **15,** 61–66.
Sokalski, I.A., Bugaev, E.V. & Klimushin, S.I., 2001. MUM: flexible precise Monte Carlo algorithm for muon propagation through thick layers of matter, *Phys. Rev. D,* **64,** 074015.
Tanaka, H. *et al.*, 2007. High resolution imaging in the inhomogeneous crust with cosmic-ray muon radiography: the density structure below the volcanic crater floor of Mt. Asama, Japan, *Earth planet. Sci. Lett.,* **263,** 104–113.
Verdu, S., 1999. Poisson communication theory, Invited talk in the International Technion Communication Day in Honor of Israel Bar-David.

APPENDIX A: DERIVATION OF MIMO CHANNEL CAPACITY FORMULA IN NOISY LIMIT

It is stated in Section 2.1 that the approach of Davis (1980) can be combined with that of Haas & Shapiro (2002) to give the result stated in eq. (9). The purpose of this appendix is not to derive the formula from first principles, but to justify how the stated result follows from these papers.

In the MIMO case, Haas and Shapiro show that a lower bound to the information capacity when using an on-off keying (OOK) approach (i.e. when data is modulated by turning the sources on and off) is given by

$$C_{\mathrm{OOK-LB}} = \sum_{m}^{M} h_m \left(\sum_{n}^{N} \gamma_{nm} A_n p^* \right), \tag{A1}$$

where the function $h_m(z)$ is defined as

$$h_m(z) = \frac{z}{R_m} \phi_m(R_m) - \phi_m(z) \tag{A2}$$

the functions ϕ_m are defined as

$$\phi_m(z) = (\Lambda_m + z) \ln(\Lambda_m + z) - \Lambda_m \ln(\Lambda_m) \tag{A3}$$

the functions R_m are defined as

$$R_m = \sum_{n}^{N} \gamma_{nm} A_n \tag{A4}$$

and p^* is a quantity which tends towards

$$p^* = \frac{1}{2} \tag{A5}$$

in the very noisy limit. The OOK case is important, as in the very noisy limit, this lower bound is equivalent to the general MIMO rate. That is

$$C_{\mathrm{OOK-LB}} = C_{\mathrm{MIMO}} \tag{A6}$$

These equations can be combined to give

$$C_{\mathrm{MIMO}} = \sum_{m}^{M} \left(\frac{1}{2} \phi_m(R_m) - \phi_m\left(\frac{1}{2} R_m \right) \right) \tag{A7}$$

In the very noisy limit, where $z \ll \Lambda_m$, the functions ϕ_m can be approximated as

$$\phi_m(z) \approx z \ln(\Lambda_m) + z + \frac{1}{2\Lambda_m} z^2 \tag{A8}$$

Using this to approximate eq. (A7) gives

$$C_{\mathrm{MIMO}} = \sum_{m}^{M} \frac{1}{8\Lambda_m} \left(\sum_{n}^{N} \gamma_{nm} A_n \right)^2 \tag{A9}$$

as stated.

APPENDIX B: ORTHOGONAL COMBINATIONS OF GEOLOGICAL QUANTITIES

In Section 2.2 it is necessary to choose a value of the matrix X_{in}, as specified in eq. (22), so that the values of ζ'_{im} satisfy eq. (23). In matrix notation, these equations can be written, respectively, as

$$\zeta' = X^{-1}\zeta \tag{B1}$$

and

$$\zeta'\zeta'^{\top} = \mathrm{diag}\,(\xi)\,, \tag{B2}$$

where it is assumed that the i index specifies row, and the m index specifies column. Substituting eq. (B1) into eq. (B2) gives

$$X^{-1}T\left(X^{-1}\right)^{\top} = \mathrm{diag}\,(\xi)\,, \tag{B3}$$

where the $N \times N$ symmetric matrix T is defined as

$$T = \zeta\zeta^{\top} \tag{B4}$$

which is equivalent to the definition given in eq. (24). If ζ contains N linearly independent columns, then it will be of full rank, and so the quantity ζx for an arbitrary non-zero vector x will also be a non-zero vector. Therefore

$$x\zeta\zeta^{\top}x^{\top} > 0 \tag{B5}$$

and so T is positive definite. A symmetric matrix can be diagonalized as

$$A = UDU^{\top}, \tag{B6}$$

where D is a diagonal matrix containing the eigenvalues of T, and U is a unitary matrix composed of the corresponding eigenvectors. As T is positive definite, these eigenvalues will be positive. Substituting eq. (B6) into eq. (B3) gives

$$X^{-1}UDU^{\top}\left(X^{-1}\right)^{\top} = \mathrm{diag}\,(\xi)\,. \tag{B7}$$

This is solved by

$$X = U \tag{B8}$$

$$D = \mathrm{diag}\,(\xi) \tag{B9}$$

which gives the result stated.

APPENDIX C: GENERALITY OF LINEAR COMBINATION TEST STATISTIC

It is assumed in Section 2.3 that a test statistic of the form

$$\Psi = \sum_m^M \gamma_m k_m \tag{C1}$$

will not greatly differ from an optimal test statistic. The expectation value of an optimal test statistic can be written in the general form

$$\psi = f\,(\lambda_1, \lambda_2, \cdots) + w, \tag{C2}$$

where f is an unknown function, and w is an arbitrary constant. The value of this constant is unimportant, as the method of analysis can be chosen so as to subtract it out. In Section 2.3, this is implicitly done by the choice of the b_{ih_n} threshold values. Taylor expanding eq. (C2) gives

$$\psi = w + f\left(\tilde{\lambda}_1, \tilde{\lambda_2}, \cdots\right) + \sum_m^M \left.\frac{\partial f}{\partial \lambda_m}\right|_{\tilde{\lambda}_m} \left(\lambda_m - \tilde{\lambda}_m\right) + \mathcal{O}\left(\left(\lambda_m - \tilde{\lambda}_m\right)^2\right), \tag{C3}$$

where the expansion points $\tilde{\lambda}_m$ are typical values for λ_m. In the low contrast case, λ_m will not greatly differ from $\tilde{\lambda}_m$ and hence the Taylor series can be truncated to first order as

$$\psi = \sum_m^M \left.\frac{\partial f}{\partial \lambda_m}\right|_{\tilde{\lambda}_m} \lambda_m, \tag{C4}$$

where w has been chosen to cancel all the other constants in this expression. Therefore the expected values of an optimal test statistic are approximately equal to a weighted sum of the expectation values of the muon counts. This suggests that the optimal test statistic is likewise similar to a weighted sum of the muon counts.

APPENDIX D: NORMAL APPROXIMATION OF PROBABILITY DISTRIBUTIONS

It is assumed in Section 2.3, that given a set of independent Poisson distributed variables k_m with expectation values λ_m, then in the case where $\sum_m^M \lambda_m \gg 1$, the probability mass function of the weighted sum

$$\Psi = \sum_m^M \gamma_m k_m \tag{D1}$$

has values which can be approximated by the probability density function of a normal distribution with mean

$$\psi = \sum_m^M \gamma_m \lambda_m \tag{D2}$$

and standard deviation σ given by

$$\sigma^2 = \sum_m^M \gamma_m^2 \lambda_m \tag{D3}$$

A weaker version of this theorem, applicable only to the special case $\lambda_m \gg 1$ can be derived by noting that the normal approximation will apply to the individual values of k_m, and hence to their sum. In the stronger theorem, however, the normal approximation need not apply to the individual values of k_m.

To proceed, it is simpler to work with the standardized quantity

$$\Psi_s = \frac{\Psi - \psi}{\sigma} \tag{D4}$$

which if the theorem is correct, should tend towards a standard normal distribution, that is a normal distribution with zero mean and unit standard deviation. The moment generating function of Ψ_s is defined as

$$\mathcal{M}\,(h) = \mathrm{E}\left[\exp\left(h\Psi_s\right)\right], \tag{D5}$$

where E denotes expectation value. This evaluates to

$$\mathcal{M}\,(h) = \exp\left(-\frac{\psi h}{\sigma} + \sum_m^M \lambda_m \left(\exp\left(\frac{\gamma_m h}{\sigma}\right) - 1\right)\right). \tag{D6}$$

Taylor expanding the innermost exponential and substituting in eq. (D3) gives

$$\mathcal{M}(h) = \exp\left(\frac{h^2}{2} + \sum_m^M \left(\frac{\lambda_m \gamma_m^3 h^3}{6\sigma^3}\right) + \mathcal{O}\left(\frac{\lambda_m}{\sigma^4}\right)\right). \tag{D7}$$

In the limit $\sum_m^M \lambda_m \gg 1$, it follows that $\sigma \gg 1$, and so eq. (D7) can be approximated as

$$\mathcal{M}(h) \approx \exp\left(\frac{h^2}{2}\right) \tag{D8}$$

which is the moment generating function of a standard normal distribution, as required.

APPENDIX E: MAXIMIZING PROBABILITY WITH RESPECT TO WEIGHTING COEFFICIENTS

It is assumed in Section 2.3, that q, as defined by

$$q(\gamma_m) = \frac{\left(\sum_m^M \tilde{\mu}_m z_m \gamma_m\right)^2}{\sum_m^M \tilde{\mu}_m \gamma_m^2} \tag{E1}$$

is maximized when $\gamma_m \propto z_m$. The point z_m can be investigated by observing how q reacts to changes from this point. Defining

$$\Delta q = q(z_m + \varepsilon_m) - q(z_m), \tag{E2}$$

where ε_m is the change from z_m, gives

$$\Delta q = \frac{\left(\sum_m^M \tilde{\mu}_m z_m (z_m + \varepsilon_m)\right)^2}{\sum_m^M \tilde{\mu}_m (z_m + \varepsilon_m)^2} - \sum_m^M \tilde{\mu}_m z_m^2 \tag{E3}$$

which can be rearranged to give

$$\Delta q = \frac{\left(\sum_m^M u_m v_m\right)^2 - \left(\sum_m^M u_m^2\right)\left(\sum_m^M v_m^2\right)}{\sum_m^M (u_m + v_m)^2}, \tag{E4}$$

where $u_m \equiv \sqrt{\tilde{\mu}_m} z_m$ and $v_m \equiv \sqrt{\tilde{\mu}_m} \varepsilon_m$. The Cauchy Schwarz inequality states that

$$\left(\sum_m^M u_m v_m\right)^2 \leq \left(\sum_m^M u_m^2\right)\left(\sum_m^M v_m^2\right) \tag{E5}$$

which together with the denominator of eq. (E4) being positive, implies that

$$\Delta q \leq 0. \tag{E6}$$

The equality in this expression corresponds to the special case where u_m is related to v_m by a common multiplier, which in turn corresponds to the fact that q is unaffected by the multiplication of γ_m by a common constant. In all other cases, Δq is negative, showing that $\gamma_m \propto z_m$ maximizes q.

APPENDIX F: LIMIT TO ANGULAR RESOLUTION

In Section 2.4, it is stated that eq. (63) gives an approximate limit to image resolution, due to the random nature of muon arrival. The situation involves variations in observed muon flux, and not the individual geological quantities that underly these variations. It is therefore necessary to take the sum of the contribution from each geological quantity. It follows from eq. (4) that $\Delta\eta_m$, the relative modification to the mean count rate for the mth measurement is given by

$$\Delta\eta_m = \sum_n^N \zeta_{nm} \Delta\chi_n. \tag{F1}$$

This can be used to simplify eq. (60), giving

$$C_\mathrm{b} = \frac{\ln(2)}{8} \sum_m^M J_m (\Delta\eta_m)^2 a_m \Delta\Omega_m. \tag{F2}$$

If the measurement for each set of directions is to be distinct, then each element of the image must be informed by at least one data bit. Therefore, for an observation time t, each term in the summation must make a contribution to the data rate of at least $1/t$, and so the condition

$$\Delta\Omega_m \geq \frac{1}{\frac{\ln(2)}{8} J_m a_m t (\Delta\eta_m)^2} \tag{F3}$$

must apply. The relation between the solid angle of a directional element, and the angle of separation from neighboring elements, is dependent on the geometric arrangement. However, it is reasonable to impose the condition $\Delta\Omega \leq \phi^2$, where ϕ is the angular separation. This is true for square grids ($\Delta\Omega = \phi^2$ in the small angle approximation), hexagonal grids ($\Delta\Omega \approx 0.866\,\phi^2$), and neighboring circles ($\Delta\Omega \approx 0.785\,\phi^2$). Consequently, the approximate condition

$$\phi_m \geq \frac{1}{\Delta\eta_m \sqrt{\frac{\ln(2)}{8} J_m a_m t}} \tag{F4}$$

applies if the properties of two different regions are to be measured independently. This condition is related to the limiting angular resolution of the detector, beyond which improvements will provide little improvement to the resulting image. The definition of this point is somewhat arbitrary, but it is sensible to choose it to be some fraction of ϕ_m. A convenient fraction is $\sqrt{\ln(2)/8} \approx 0.294$, which is chosen so as to remove numerical factors from the resulting equation. Choosing this fraction gives the result stated by eq. (63).

APPENDIX G: MUON ENERGY SAMPLING

The simplest approach for calculating the survival probability function is to initialize the simulation with random energies chosen from the probability density function of muon energy at ground level. The survival probability Ξ, as a function of distance s can then be estimated as

$$\Xi(s) = 1 - \frac{1}{N} \sum_{n=1}^N \{s_n \leq s\}, \tag{G1}$$

where s_n is the distance attained by the nth simulation out of N. The curly brackets denote a value of 1 if the condition within is true, and 0 otherwise.

This approach is problematic for several reasons. At large (multi kilometre) propagation distances, the surviving muons are usually those which started with uncommonly high energy. When only a few of these muons are sampled, there will be large random error. Another problem is that the ground level energy distributions are dependent on direction, and so a separate set of simulations would be required for each zenith angle.

An alternative approach is to calculate muon survival functions for fixed energies, and then to combine these with the initial energy spectrum to produce an overall survival function (Kudryavtsev

2009). This, however, requires a two dimensional set of simulations, as a set of simulations must be performed for each individual energy value.

Instead, the simulations were initialized with randomly chosen energies, but from a distribution which overrepresented the high energy values. A power law distribution of the form

$$\mathrm{P}(E) = \frac{\alpha - 1}{E_0} \begin{cases} 0 & ; E < E_0 \\ \left(\frac{E}{E_0}\right)^{-\alpha} & ; E \geq E_0 \end{cases} \tag{G2}$$

was used, which is constructed so as to exclude muons below the energy cutoff at $E_0 = 100$ GeV. This is only valid when $\alpha > 1$, as for lower values, power law integrals are divergent, and hence the distribution cannot be normalized.

Random numbers conforming to this distribution can be generated using inverse transform sampling. A random value for E can be obtained by generating a random value x following a continuous uniform distribution between 0 and 1, and then using the formula

$$E = E_0\, x^{\frac{1}{\alpha - 1}} \tag{G3}$$

to generate a value following the power law distribution. To calculate the survival probability function, eq. (G1) was modified to give

$$\Xi(s, \theta) = 1 - \frac{\sum_{n=1}^{N} p(E_n, \theta)\{s_n \leq s\}}{\sum_{n=1}^{N} p(E_n, \theta)}, \tag{G4}$$

where E_n is the initial energy for the nth simulation, and $p(E_n, \theta)$ is a correction for the over representation of higher energies. It is given by

$$p(E, \theta) = g(E, \theta) \left(\frac{E}{E_0}\right)^{\alpha}, \tag{G5}$$

where $g(E, \theta)$ is the real probability density function of muon energy at ground level. eq. (G4) was used to produce the survival probability functions shown in Fig. 4. A power law value of $\alpha = 1.5$ was chosen, to give good representation to the higher energies, while retaining normalizability.

APPENDIX H: IMPLEMENTATION OF RAY TRACING

To perform muon rayt racing, it is first necessary to parametrize directions of arrival in a convenient form. This is done by back-projecting the directions onto a horizontal plane at ground level. Such a projection (which is equivalent to gnomonic projection in map-making) is used for the images shown in Section 4. A small region on the image plane with area ΔA_m will cover a solid angle of

$$\Delta\Omega_m = \frac{\cos(\theta_m)}{s_m^2}\, \Delta A_m, \tag{H1}$$

where s_m is the distance from region to the detector, and θ_m is zenithal angle. The image plane can then be rasterized, where each pixel corresponds to a ray tracing path and represents an element of solid angle given, as by eq. (H1).

Once a set of paths have been constructed, each path is passed to a ray tracing function which determines the regions of the model that the path intersects, and calculates the path length of each intersection. The ray tracer output is then used to calculate effective distance for each path, as given by eq. (67). Putting these into eq. (71) then gives the muon flux for each pixel.

The model used in Section 4 is represented using corner-point geometry, a grid system which is widely used in hydrocarbon reservoir simulations (Ponting 1989). This is similar to a Cartesian grid system (which remains a special case), but instead of the cells being cuboids, they can be distorted into irregular hexahedra. This is done by constructing a set of $x \times y$ cell columns, where the edges at which each column meets its neighbours (together with the outer boundary of the grid) are specified by a set of $(x + 1) \times (y + 1)$ quasi-vertical reference lines. The individual cells within each column are then defined by specifying their corners in terms of the distances along the reference lines. As the cell faces have corners that are not necessarily coplanar, the faces are in fact constructed from four triangles (one for each edge) which join together at the centroid of the corners (i.e. the mean of their Cartesian coordinates). Therefore, each cell may in fact have more than six faces, albeit while retaining an approximately hexahedral shape.

The corner-point system is designed to naturally accommodate uneven geological layering and fault lines, while retaining the convenience of storing the geological data in ordinary 3-D arrays. The shape of the geological layers are specified by the shape of the grid system, and so the grid geometry is just as much a part of the geological model as the data within the 3-D arrays. In fact, the modelling used in Section 4 was done not by assigning values to each cell in a grid, but by taking an array of known values for the substances involved, and then shaping the corresponding grid system such that their boundaries matched that of the geology.

APPENDIX I: MODEL OF CARBON DIOXIDE MOVEMENT

To describe the subsurface movement of the injected CO_2, an analytical model based on that of Mathias *et al.* (2009) was used.

The model assumes that a set fraction of the Bunter sandstone consists of a brine filled pore space, with the remainder being solid. The porosity ϕ is defined as the fraction (by volume) that the pore space occupies. The structure of the rock is assumed to vary over length scales far shorter than the resolution of the muon images, so that bulk averages of the material properties can be used. Therefore, prior to CO_2 injection, the bulk density of the rock is given by

$$\rho_{\mathrm{unfilled}} = (1 - \phi)\, \rho_{\mathrm{s}} + \phi \rho_{\mathrm{w}}, \tag{I1}$$

where ρ_{w} and ρ_{s} are the intrinsic phase averaged densities of the brine and solid rock, respectively. When supercritical CO_2 is added, it displaces some of the brine, but a fraction is assumed to be immobile. Therefore, the bulk density of rock considered filled to its maximum extent with CO_2 is given by

$$\rho_{\mathrm{filled}} = (1 - \phi)\, \rho_{\mathrm{s}} + \phi S_{\mathrm{rw}} \rho_{\mathrm{w}} + \phi (1 - S_{\mathrm{rw}})\, \rho_{\mathrm{n}}, \tag{I2}$$

where ρ_{n} is the density of the supercritical CO_2, and S_{rw} is the fraction of brine assumed to be immobile. (The brine and the supercritical CO_2 are in some sources referred to as the 'wetting' and 'non-wetting' phases, hence the subscripts 'w' and 'n'. Similarly, the subscript 'rw' stands for residual wetting.) The densities are assumed to be constants. This requires that they be calculated for some reference temperature and pressure, with variations in these properties assumed to have negligible effect. The multiphase components are assumed to be immiscible and non-reactive. The rock porosity and the residual wetting are also assumed to be constants. The wetting and non-wetting components have relative permeabilities k_{rw} and k_{rn}, and fluid viscosities μ_{n} and μ_{w}, which are also assumed to be constants.

It is assumed that there is a well defined fluid interface between the filled and unfilled regions. The filled region, being less dense, floats on top, reaching to the top of the Bunter layer. The purpose of the flow model is therefore to calculate the shape of the surface separating the filled and unfilled regions. This flow is assumed to be largely driven by the increased fluid pressure at the injection point, which represents a vertical injection well. It is also assumed that the flow is non-inertial, the pore system is undeformable, capillary pressure is negligible, and that the horizontal boundaries of the Bunter layer are too distant to have any effect on fluid movement.

If CO_2 is injected at a rate M_n along a vertical injection well penetrating through a Bunter layer of height H, then an approximate analytical solution for the plume thickness h, as a function of time t and radial horizontal distance from the injection point x can be derived from Mathias *et al.* (2009) as

$$h(x,t) \approx \begin{cases} H & ; x \leq \gamma\theta\sqrt{t} \\ \dfrac{H\gamma\theta}{1-\gamma}\dfrac{\sqrt{t}}{x} - \dfrac{H\gamma}{1-\gamma} & ; \gamma\theta\sqrt{t} < x < \theta\sqrt{t} \\ 0 & ; \theta\sqrt{t} \leq x \end{cases} \tag{I3}$$

where the constant terms γ and θ are given by

$$\gamma = \frac{\mu_n k_{rw}}{k_{rn}\mu_w} \tag{I4}$$

$$\theta = \sqrt{\frac{M_n}{\pi H \phi (1 - S_{rw}) \rho_n \gamma}} \tag{I5}$$

Geometrically, this solution describes a cylindrical central region [the first domain in eq. (I3)] surrounded by an outer hyperboloidal region [the second domain in eq. (I3)] which extends to a radius of $\theta\sqrt{t}$ before reaching a value of zero [the final domain in eq. (I3)]. Both horizontal dimensions scale with time as $\sqrt{t}$. The total volume is therefore proportional to t, and hence proportional to the mass of CO_2 injected. The applicability of these equations has been explored by Mathias *et al.* (2009, 2011b), Nordbotten *et al.* (2005) and Nordbotten & Celia (2006).

Two further assumptions are made to adapt this model to the real geology. First, to account for the fact that the boundary between the Bunter sandstone and the layer above is uneven, the depth of the fluid interface between the filled and unfilled regions is altered so that the thickness, h, of the CO_2 plume conforms to eq. (I3). Secondly, to account for the fact that the well does not penetrate to the base of the Bunter sandstone (as the derivation of eq. (I3) assumes) an impermeable layer is assumed to exist beneath the base of the well. This is partially justified by the fact that the supercritical CO_2 is less dense than the brine and hence would not sink, but in reality, the injection pressure will drive the boundary between the filled and unfilled regions slightly below the bottom of the well.